Impact of Variation in Vector Space on the performance of Machine and Deep Learning Models on an Out-of-Distribution malware attack Detection

1st Tosin Ige
Department of Computer Science
The University of Texas at El Paso
Texas, USA
toige@miners.utep.edu

2nd Christopher Kiekintveld
Department of Computer Science
The University of Texas at El Paso
Texas, USA
cdkiekintveld@utep.edu

3rd Aritran Piplai
Department of Computer Science
The University of Texas at El Paso
Texas, USA
apiplai@utep.edu

4th Asif Rahman
dept. name of Computer Science
The University of Texas at El Paso
Texas, USA
arahman3@miners.utep.edu

5th Olukunle Kolade
Department of Computer Science
University of North Carolina
North Carolina, USA
ookol@unc.edu

***Abstract*—Several state-of-the-art machine and deep learning models in the mode of adversarial training, input transformation, self adaptive training, adversarial purification, zero-shot, one-shot, and few-shot meta learning had been proposed as a possible solution to an out-of-distribution problems by applying them to wide arrays of benchmark dataset across different research domains with varying degrees of performances, but investigating their performance on previously unseen out-of-distribution malware attack remains elusive. Having evaluated the poor performances of these state-of-the-art approaches in our previous research on an out-of-distribution attack. In this research, we dived deeper to understand why they works better for other domain dataset but with poor performance on available benchmark malware dataset like Malimg, Malevis, Sorel, and Avast CTU malware dataset. We explored the both the embedding and vector spaces in datasets and compare them with that from other research domain, and find a surprising wide variation between the embedding and vector spaces in malware datasets. We assert that current state-of-the-art machine and deep learning models does not address the wide variation of embedding and vector spaces which are peculiar to malware dataset, hence their poor performance on out-of-distribution attack classification, and so concluded that addressing this variation in embedding and vector spaces will bring about substantial increase in detection of previously unseen out-of-distribution attack**

***Index Terms*—Malware, Malware Attack, Machine Learning, Deep Learning, Out-of-Distribution attack**

I. Introduction

Cybercriminals often use all forms of malicious software called malware for several purposes such as deception of inducing potential victim to divulge financial information personal details for identity theft or hijacking several computers to launch distributed denial-of-service attack against network of computer [16], [30]. The potency of malware to successfully infiltrate any system no matter how sophisticated made it an indispensable tool available to cybercriminals today, as malware had proven to be highly successful in the extraction of sensitive data which could be used by cybercriminals against their victim. Several approaches had been widely proposed and adopted to combat the rampant threat of malware attack among which machine learning (ML) and Deep Learning (DL) had been the most promising [33] but the out-of-distribution (OOD) problems had lead to vulnerabilities of machine and deep learning based approaches against previously unseen malware family or new variant of an existing family. This is due to the fact that current state-of-the-art approaches are based on the assumption that identically and independently distributed (IID) data will be available in test time [6], [36]–[41], which are unfortunately not true in new world scenarios [2]. Hence, the close-world assumption of identically and independently distributed are violated whenever state-of-the-art machine or deep learning based model are deploy in real-world scenarios in the presence of previously unseen out-of-distribution mal- ware family or variants of an existing family, the high failure rate of state-of-the-art approaches to previously unseen OOD malware is cyberattack [9]–[11], [14].

This is evident by the established fact of malware being the fastest-growing threat with 41% of enterprises witnessing a malware attack in just concluded year 2023 followed by phishing and ransomware attack. In year 2023 alone, the number of enterprises experiencing ransomware attacks increased by over 27% with only 8% of businesses attacked resorting to paying the ransom demands resulting in significant financial loss in addition to losses incurred due to downtime. There are 95 new families of malware in year 2022 alone averaging 1 new family every 4 days aside variants while year 2023

979-8-3503-6491-0/24/$31.00 ©2024 IEEE

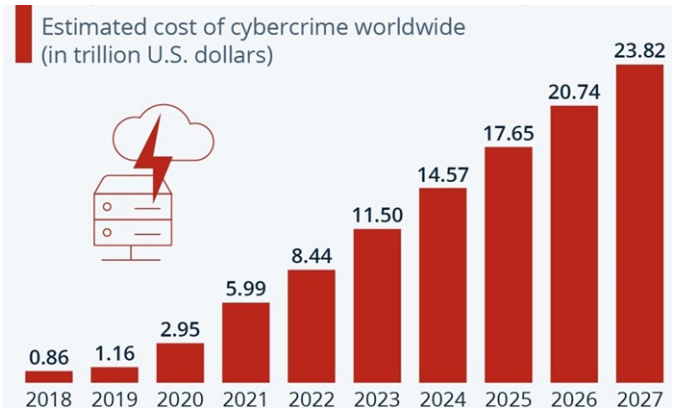

Fig. 1. Cybercrime expected to Skyrocket in coming years
Sources: Statista Technology Market Outlook, National Cybersecurity Organizations, FBI, IMF

witness 43 new malware families averaging 1 new malware family per week aside variants making emerging malware families a major threat to cybersecurity causing damages worth billions of Dollars annually [29], [31]. The ease with which attacker creates new variants of malware coupled with the rate at which new variants are being release poses a real challenge both for their detection, identification and classification, reason being that machine learning and deep learning model are only effective in detecting previously seen variants during training [3].

The major cause of this problem is the availability of sophisticated tools for cyberattackers to rapidly create an out-of-distribution variants of existing malware family or new malware family against which State-of-the-art (SOTA) machine learning and deep learning models remain vulnerable. Hence, Statista Technology Market Outlook, National Cyber Security Organizations, FBI, and IMF had projected an increase in cybercrime world wide 1 [8], [12], [13] [42]–[46].

While several SOTA approaches had been proposed to address the problem of OOD in both ML and DL subfield of artificial intelligence (AI) on several benchmark datasets, none of the SOTA approaches had been applied to OOD malware dataset thereby leaving a gap to filled. We started by applying and training SOTA models on 4 different benchmark malware dataset (Sorel,Malevis,Malimg, and Avast), after seeing the poor performance of the current state-of-the-art deep learning models and approaches on an OOD malware dataset, we proceeded to investigate the possible cause by converting each variants in each malware family to bytes and calculating the mean square error (MSE) of each family member, same procedure was also also repeated for other dataset on which SOTA models have good performance. Our result shows wide variation between variants from the same malware family through the Mean Square Error (MSE) and vector space while other datasets shows little to no variation between samples of same class. Hence, unlike other dataset where sample from a given class can represent the whole class, malware sample from a given family does not give a true representation of the family due to wide variation between samples from same family which we measured in form of the Mean Square Error, current SOTA OOD techniques does not give provision for this wide MSE spread among samples from the same malware family, hence, the reason for their poor performance on OOD malware. Our research here has two main objectives;

- First aim is to investigate how variation of variants from same malware family leads to poor performance of state-of-the-art machine and deep learning model on an previously unseen out-of-distribution attack while performing better on other domain outside cybersecurity. This is crucial because understanding this will enable development of new state-of-the-art models for effective out-of-distribution malware attack detection, while also enhancing current state-of-the-art machine and deep learning models.
- Our second objective is to propose base on the result. Hence, we assert future research direction will center around mitigating the impact of the variation in vector spaces. so, we proposed that future state-of-the-art machine and deep learning models will have to address this limitation by, (1) exploitation of the in-dimensional embedding space between malware variants from the same malware family to account for all variations (2) exploitation of the inter-dimensional space from different malware family and (3) real time dynamic adjustment of data points

II. Related Work

Among the several types of cyber threat, malware attacks remains the top threat defying recent advances due to the ease and potency of creating new variants from existing malware. In our research, we emphasizes on the out-of-distribution problem owing to the fact that each and every current state-of-the-art approaches performs very poorly against an out-of-distribution malware. Subsequently, we exploit the spaces between each and every samples from the same malware family in our proposed deep learning-based framework for effective classification of previously unseen out-of-distribution malware attack.In the following sections, we present related work based on the aforementioned research question, providing a comprehensive overview of the existing literature and relevant findings.

III. Malware Obfuscation and Behavioural Analysis

Over the years, several methods had been proposed for effective detection of malware which are broadly classified into Static and dynamic categories. Static malware detection methods such as deployed in [18], [24], [27], [32] are rule based and heavily relies header information, file hashes and Opcodes features to detect malware, the problem static method is that they can be evaded by polymorphism and obfuscation techniques [1], [20], [21]. On the other hand, dynamic malware detection method uses behavioral-based features such as the

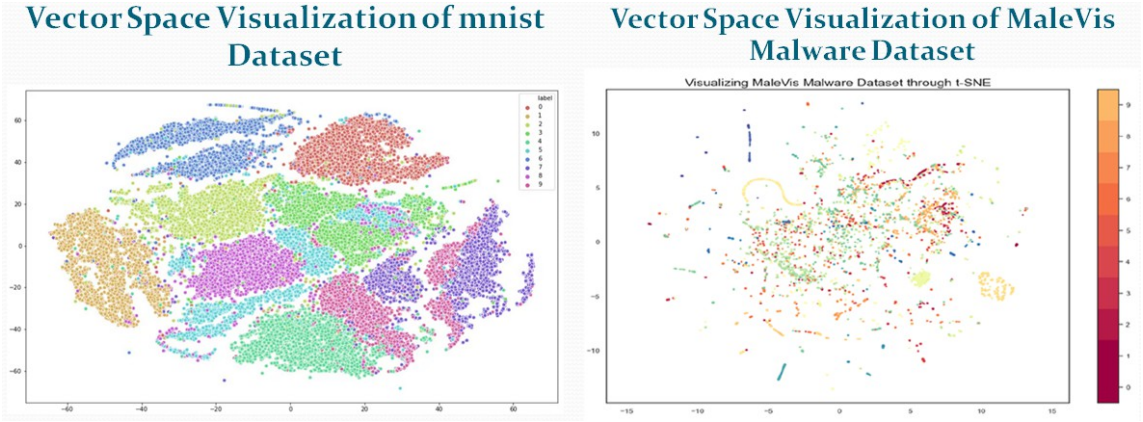

Fig. 2. t-sne Side by side comparison of variation in vector spaces between variants of same malware family (MaleVis) Malware dataset compare with Mnist dataset showing wide and overlapping vector spaces between variants of same malware family

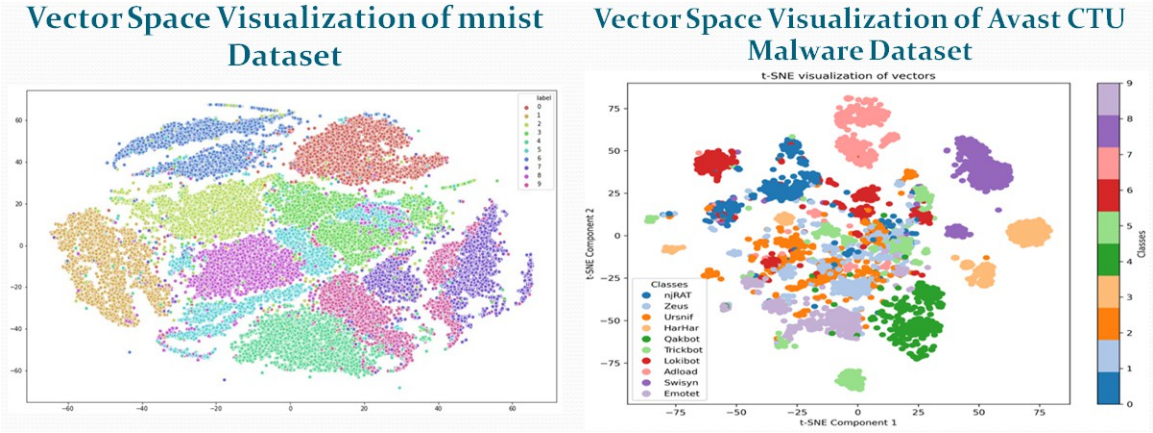

Fig. 3. t-sne Side by side comparison of variation in vector spaces between variants of same malware family (Avast CTU) Malware dataset compare with Mnist dataset showing wide and overlapping vector spaces between variants of same malware family

monitoring of the process-level behavior of malware to classify it [15].

But according to Huan Zhang [34], behavioral-based process-level detection malware methods have vulnerabilities due to the susceptibilities to evasive tactics such as multi-process techniques and junk code injection since a malware like ransomware can easily have multiple child processes with each process executing small portion of the overall task while at the same time evading detection by mimicking benign behavior [35].

One obvious observation in all the machine and deep learning state-of-the-art approaches to malware classification as seen in **??** is they are not train for novel variant classification, hence their vulnerabilities to previously unseen or novel out-of-distribution malware, and so will surely falter when attack with more recent sophisticated malware. My proposed deep learning based framework will address this gap to make a significant contribution to both field of artificial intelligence and cybersecurity.

Certain problems like generalization, convergence and divergence are peculiar with few-shot learning considering that it is aimed at categorizing the new classes of previously unseen samples in the training set having been given only few samples of from each class, Over the years several new algorithms and adjustment to the current state-of-the-art algorithm have been developed in order to address some of the more peculiar problem associated with few-shot learning [22] such as the adoption of Probabilistic models based on Bayesian learning [4], [5], Generative models with probability density functionality [17], [19], image transformation [7], [23], Using memory augmented in neural networks [26] , Meta learning [17], [25] and Metric learning [28].

IV. Research Methodology

1) Stage 2 - Investigating the Uniqueness of Malware dataset and Identification of in-Distributional Dimensional Space : In order to have a thorough analysis of the underlying structure of malware, we use holistic approach whereby each malware family was treated as separate entity. For each variant member of a malware family, each of the variants from each malware family were converted to a one-channel image for proper storage after which each pixels were converted to a NumPy array and save. Each of the saved NumPy array were

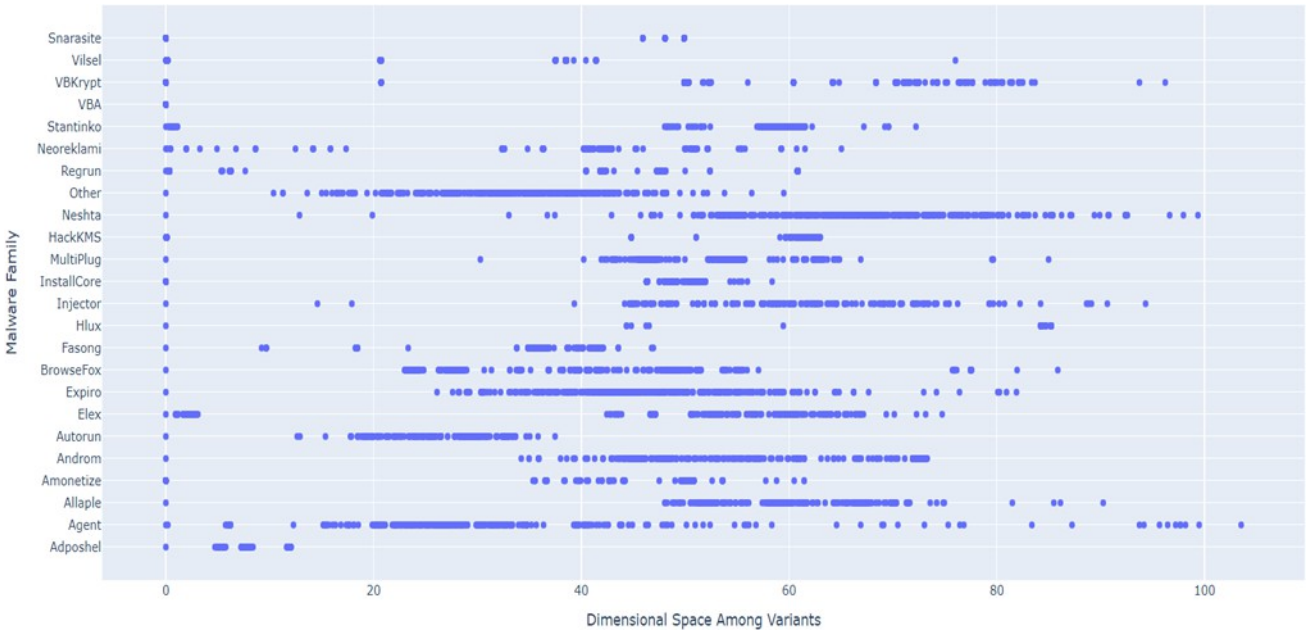

Fig. 4. Visualization of the distribution of Mean Square Error(MSE) between variants from same malware family showing wider MSE

converted back to an image and compared with the original image both by physical inspection and byte array to ensure that each of the generated arrays is a true representation of the original malware image. On confirmation that each of the saved NumPy array is a true representation of the original image, the following mode of visualization was adopted;

A. t-SNE for vector space representation of each pixels

It becomes clear that there is an existence of variation in dimensional space between each variants of every malware family which had not been previously exploited by any of the existing state-of-the-art machine and deep learning based approaches for an out-of-distribution detection, unlike other dataset where a sample can give true representation of that class, a single malware variant does not actually give any true representation of that family. We conclude that this variation in dimensional space between family of malware had not been exploited and it is the reason why all existing state-of-the-art OOD approaches performed poorly when applied to malware in an out-of-distribution settings 32.

B. MSE and scatter mapping for proper exploration of the wideness in the Mean Square Error (MSE)

It becomes clear that there is an existence of variation in dimensional space between each variants of every malware family which had not been previously exploited by any of the existing state-of-the-art machine and deep learning based approaches for an out-of-distribution detection, unlike other dataset where a sample can give true representation of that class, a single malware variant does not actually give any true representation of that family. We conclude that this variation in dimensional space between family of malware had not been exploited and it is the reason why all existing state-of-the-art OOD approaches performed poorly when applied to malware in an out-of-distribution settings 4.

V. Conclusion

In this research, we investigate how variation in variants of same malware family leads to poor performance of state-of-the-art machine and deep learning model on an previously unseen out-of-distribution malware attack while performing better on other domain outside cybersecurity, and diving deeper to understand the poor performances on available benchmark malware like Malimg, Malevis, Sorel, and Avast CTU malware dataset. We explored both the embedding and vector spaces in malware datasets and compare them with that from other research domain, and find a surprising wide variation between the embedding and vector spaces in malware datasets even among varints from same malware family. We assert that current state-of-the-art machine and deep learning models does not address the wide variation of embedding and vector spaces which are peculiar to malware dataset, hence their poor performance on out-of-distribution attack classification, and so concluded that addressing this variation in embedding and vector spaces will bring about substantial increase in detection of previously unseen out-of-distribution attack.

Considering the impact of variation in vector and embedding spaces on the poor performance of current state-of-the-art models on the detection of previously unseen out-of-distribution malware attack, future research direction will center around mitigating the impact of these vector spaces. Hence, we proposed that future state-of-the-art machine and deep learning models will have to address this limitation by, (1) exploitation of the in-dimensional embedding space

between malware variants from the same malware family to account for all variations (2) and exploitation of the inter-dimensional space from different malware family.

REFERENCES

[1] Fatimah Aldauiji, Omar Batarfi, and Manal Bayousef. Utilizing cyber threat hunting techniques to find ransomware attacks: A survey of the state of the art. *IEEE Access*, 10:61695–61706, 2022.

[2] Murat Dundar, Balaji Krishnapuram, Jinbo Bi, and R Bharat Rao. Learning classifiers when the training data is not iid. In *IJCAI*, volume 2007, pages 756–61. Citeseer, 2007.

[3] Manuel Egele, Theodoor Scholte, Engin Kirda, and Christopher Kruegel. A survey on automated dynamic malware-analysis techniques and tools. *ACM computing surveys (CSUR)*, 44(2):1–42, 2008.

[4] Li Fe-Fei et al. A bayesian approach to unsupervised one-shot learning of object categories. In *proceedings ninth IEEE international conference on computer vision*, pages 1134–1141. IEEE, 2003.

[5] Li Fei-Fei, Robert Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence*, 28(4):594–611, 2006.

[6] Navid Ghassemi and Ehsan Fazl-Ersi. A comprehensive review of trends, applications and challenges in out-of-distribution detection. *arXiv preprint arXiv:2209.12935*, 2022.

[7] Bharath Hariharan and Ross Girshick. Low-shot visual object recognition. *arXiv preprint arXiv:1606.02819*, 2(5), 2016.

[8] Tosin Ige. Exploiting the in-distribution embedding space with deep learning and bayesian inference for detection and classification of an out-of-distribution malware, 2024.

[9] Tosin Ige and Christopher Kiekintveld. Performance comparison and implementation of bayesian variants for network intrusion detection. *arXiv preprint arXiv:2308.11834*, 2023.

[10] Tosin Ige, Christopher Kiekintveld, and Aritran Piplai. Deep learning-based speech and vision synthesis to improve phishing attack detection through a multi-layer adaptive framework. *arXiv preprint arXiv:2402.17249*, 2024.

[11] Tosin Ige, Christopher Kiekintveld, and Aritran Piplai. An investigation into the performances of the state-of-the-art machine learning approaches for various cyber-attack detection: A survey. *arXiv preprint arXiv:2402.17045*, 2024.

[12] Tosin Ige, Christopher Kiekintveld, Aritran Piplai, Amy Wagler, Olukunle Kolade, and Bolanle Matti. Towards an in-depth evaluation of the performance, suitability and plausibility of few-shot meta transfer learning on an unknown out-of-distribution cyber-attack detection, 2024.

[13] Tosin Ige, Christopher Kiekintveld, Aritran Piplai, Amy Wagler, Olukunle Kolade, and Bolanle Hafiz Matti. An in-depth investigation into the performance of state-of-the-art zero-shot, single-shot, and few-shot learning approaches on an out-of-distribution zero-day malware attack detection, 2024.

[14] Tosin Ige, William Marfo, Justin Tonkinson, Sikiru Adewale, and Bolanle Hafiz Matti. Adversarial sampling for fairness testing in deep neural network. *arXiv preprint arXiv:2303.02874*, 2023.

[15] Amin Kharaz, Sajjad Arshad, Collin Mulliner, William Robertson, and Engin Kirda. {UNVEIL}: A {Large-Scale}, automated approach to detecting ransomware. In *25th USENIX security symposium (USENIX Security 16)*, pages 757–772, 2016.

[16] Seung Hyun Kim, Qiu-Hong Wang, and Johannes B Ullrich. A comparative study of cyberattacks. *Communications of the ACM*, 55(3):66–73, 2012.

[17] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-sgd: Learning to learn quickly for few-shot learning. *arXiv preprint arXiv:1707.09835*, 2017.

[18] May Medhat, Samir Gaber, and Nashwa Abdelbaki. A new static-based framework for ransomware detection. In *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, pages 710–715. IEEE, 2018.

[19] Akshay Mehrotra and Ambedkar Dukkipati. Generative adversarial residual pairwise networks for one shot learning. *arXiv preprint arXiv:1703.08033*, 2017.

[20] Andreas Moser, Christopher Kruegel, and Engin Kirda. Limits of static analysis for malware detection. In *Twenty-third annual computer security applications conference (ACSAC 2007)*, pages 421–430. IEEE, 2007.

[21] Harun Oz, Ahmet Aris, Albert Levi, and A Selcuk Uluagac. A survey on ransomware: Evolution, taxonomy, and defense solutions. *ACM Computing Surveys (CSUR)*, 54(11s):1–37, 2022.

[22] N O'Mahony, Sean Campbell, Anderson Carvalho, L Krpalkova, Gustavo Velasco Hernandez, Suman Harapanahalli, D Riordan, and J Walsh. One-shot learning for custom identification tasks; a review. *Procedia Manufacturing*, 38:186–193, 2019.

[23] Frederik Pahde, Mihai Puscas, Jannik Wolff, Tassilo Klein, Nicu Sebe, and Moin Nabi. Low-shot learning from imaginary 3d model. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 978–985. IEEE, 2019.

[24] Subash Poudyal, Kul Prasad Subedi, and Dipankar Dasgupta. A framework for analyzing ransomware using machine learning. In *2018 IEEE symposium series on computational intelligence (SSCI)*, pages 1692–1699. IEEE, 2018.

[25] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *International conference on learning representations*, 2016.

[26] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-learning with memory-augmented neural networks. In *International conference on machine learning*, pages 1842–1850. PMLR, 2016.

[27] Shina Sheen and Ashwitha Yadav. Ransomware detection by mining api call usage. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 983–987. IEEE, 2018.

[28] Yisheng Song, Ting Wang, Puyu Cai, Subrota K Mondal, and Jyoti Prakash Sahoo. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Computing Surveys*, 55(13s):1–40, 2023.

[29] K Muthamil Sudar, P Deepalakshmi, P Nagaraj, and V Muneeswaran. Analysis of cyberattacks and its detection mechanisms. In *2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, pages 12–16. IEEE, 2020.

[30] Nida Tariq. Impact of cyberattacks on financial institutions. *Journal of Internet Banking and Commerce*, 23(2):1–11, 2018.

[31] Daniele Ucci, Leonardo Aniello, and Roberto Baldoni. Survey of machine learning techniques for malware analysis. *Computers & Security*, 81:123–147, 2019.

[32] Aldin Vehabovic, Hadi Zanddizari, Nasir Ghani, Farooq Shaikh, Elias Bou-Harb, M Safaei Pour, and Jorge Crichigno. Data-centric machine learning approach for early ransomware detection and attribution. In *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, pages 1–6. IEEE, 2023.

[33] Aston Zhang, Zachary C Lipton, Mu Li, and Alexander J Smola. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021.

[34] Huan Zhang, Lixin Zhao, Aimin Yu, Lijun Cai, and Dan Meng. Ranker: Early ransomware detection through kernel-level behavioral analysis. *IEEE Transactions on Information Forensics and Security*, 2024.

[35] Chijin Zhou, Lihua Guo, Yiwei Hou, Zhenya Ma, Quan Zhang, Mingzhe Wang, Zhe Liu, and Yu Jiang. Limits of i/o based ransomware detection: An imitation based attack. In *2023 IEEE Symposium on Security and Privacy (SP)*, pages 2584–2601. IEEE, 2023.

[36] Ige, T., Marfo, W., Tonkinson, J., Adewale, S., & Matti, B. H. (2023). Adversarial sampling for fairness testing in deep neural network. arXiv preprint arXiv:2303.02874.

[37] Ige, Tosin, Christopher Kiekintveld, Aritran Piplai, Amy Wagler, Olukunle Kolade, and Bolanle Hafiz Matti. "An in-Depth Investigation Into the Performance of State-of-the-Art Zero-Shot, Single-Shot, and Few-Shot Learning Approaches on an Out-of-Distribution Zero-Day Malware Attack Detection." In 2024 International Symposium on Networks, Computers and Communications (ISNCC), pp. 1-6. IEEE, 2024.

[38] Ige, Tosin, Christopher Kiekintveld, Aritran Piplai, Amy Wagler, Olukunle Kolade, and Bolanle Hafiz Matti. "Towards an in-Depth Evaluation of the Performance, Suitability and Plausibility of Few-Shot Meta Transfer Learning on An Unknown Out-of-Distribution Cyber-attack Detection." In 2024 International Symposium on Networks, Computers and Communications (ISNCC), pp. 1-6. IEEE, 2024.

[39] Ige, T., Kiekintveld, C., & Piplai, A. (2024, May). An investigation into the performances of the state-of-the-art machine learning approaches for various cyber-attack detection: A survey. In 2024 IEEE International Conference on Electro Information Technology (eIT) (pp. 135-144). IEEE.

[40] Ige, T., Kiekintveld, C., & Piplai, A. (2024). Deep Learning-Based Speech and Vision Synthesis to Improve Phishing Attack Detection through a Multi-layer Adaptive Framework. arXiv preprint arXiv:2402.17249.

[41] Ogaga, D. and Abiodun Olalere. 2023 "Evaluation and Comparison of SVM, Deep Learning, and Naïve Bayes Performances for Natural Language Processing Text Classification Task" Preprints. https://doi.org/10.20944/preprints202311.1462.v1

[42] Abiodun Olalere , “Impact of Data Warehouse on Organization Development and Decision making (A Case study of United Bank for Africa and Watchlocker PLC) ” International Journal of Research and Scientific Innovation (IJRSI) vol.10 issue 1, pp.36-45 January 2023 URL: https://www.rsisinternational.org/journals/ijrsi/digital-library/volume-10-issue-1/36-45.pdf

[43] Agboro, D. The Use of Machine Learning Methods for Image Classification in Medical Data. URL: https://philpapers.org/rec/AGBTUO

[44] Ogaga, Destiny and Zhao, Haoning, The Rise of Artificial Intelligence and Machine Learning in HealthCare Industry (May 15, 2023). International Journal of Research and Innovation in Applied Science , Available at SSRN: https://ssrn.com/abstract=4483867

[45] Destiny Ogaga, Haoning Zhao “The Rise of Artificial Intelligence and Machine Learning in HealthCare Industry ” International Journal of Research and Innovation in Applied Science (IJRIAS) volume-8-issue-4, pp.250-253 April 2023 DOI: https://doi.org/10.51584/IJRIAS.2023.8426

[46] Ogaga, Destiny. "COURSE REGISTRATION AND EXAM PROCESSING SYSTEM." URL: https://www.researchgate.net/publication/374725473_COURSE_REGISTRATION_AND_EXAM_PROCESSING_SYSTEM